ORIGINAL MANUSCRIPT

emotion review

Emotion Review
Vol. 0, No. 0 (February 2022) 1–12

ISSN:1754-0739
DOI: 10.1177/17540739221082203
https://journals.sagepub.com/home/emr

What is an Emotion? A Connectionist Perspective

Gaurav Suri[1] and James J. Gross[2]
[1]Department of Psychology, San Francisco State University, San Francisco, CA, USA
[2]Department of Psychology, Stanford University, Stanford, CA, USA

Abstract

Researchers often disagree as to whether emotions are largely consistent across people and over time, or whether they are variable. They also disagree as to whether emotions are initiated by appraisals, or whether they may be initiated in diverse ways. We draw upon Parallel-Distributed-Processing to offer an algorithmic account in which features of an emotion instance are bi-directionally connected to each other via conjunction units. We propose that such indirect connections may be innate as well as learned. These ideas lead to the development of the Interactive Activation and Competition framework for Emotion (IAC-E) which allows us to specify when emotions are consistent and when they are variable, as well as when they are appraisal-led and when they are input-agnostic.

Keywords
computational models, emotion theories, parallel distributed processing, neural networks

Well over a century ago, William James famously asked "What is an emotion?" (James, 1884). Over the ensuing decades, a high-level consensus has emerged that an emotion may be seen as a response to an event that has perceived significance for the individual (Keltner & Gross, 1999). Such responses typically feature relatively short-lived changes in the domains of behavior, experience, and physiology, among others (Mauss et al., 2005).

Beyond this high-level agreement, however, consensus about the nature of emotion has proven to be stubbornly elusive (Griffiths, 2004; Moors, 2017). Two points of particular disagreement revolve around two fundamental questions about emotion: First, to what extent are emotion-related responses (to similar situations) consistent across time and across people? Second, do emotion-related responses generally unfold serially over time (with one type of response likely to occur before another type of response), or are they parallel processes in which various response types interact with each other?

Are Emotions Generally Consistent or Variable?

Consider two instances of "fear" – either in the same person at two different times, or in two different people. Some researchers have noted that such instances often feature consistent responses – such as a widening of the eyes or an increased pulse rate – that are characteristic of many instances of fear (Fullana et al., 2016; Öhman & Wiens, 2003). More generally, researchers with basic emotion perspectives (Cowen & Keltner, 2017; Ekman, 1992; Ekman & Cordaro, 2011; Izard, 2007; Lench et al., 2011; Panksepp, 2004, 2007; Tracy & Randles, 2011; Tracy & Robins, 2004) have emphasized the ubiquitous consistency in response patterns in emotions of the same type across time and across people.

Other researchers have noted that experience and context often create variability in emotional responses. According to this view, two instances of "fear" may be dissimilar from one another due to differences in the specific contexts that produce these emotions (e.g. Crivelli et al., 2016; Russell, 1991). For example, some instances of fear may involve freezing in place, whereas others may involve rapid flight. More generally, researchers with constructed emotion perspectives (Averill, 1983; Barrett, 2006; Lindquist, 2013; Lindquist & Barrett, 2008; Mesquita & Frijda, 2016; Russell, 2003, 2005, 2009) have emphasized the ubiquitous variability of emotion responses. This variability is thought to arise from the influence of context-specific variables and prior-experience.

Corresponding author: Gaurav Suri, Department of Psychology, San Francisco State University, San Francisco, California, USA.
Email: suri@sfsu.edu

Many basic emotion and constructed emotion researchers refrain from endorsing the most-extreme, absolute version of their respective perspectives. For example, some basic emotion researchers have noted that emotion-instances are distinct from emotion-schemas which refer to the dynamic interplay of basic emotions and higher order cognition (Izard, 2007). This distinction allows for learning-based effects to cause variances in otherwise consistent emotion instances. Similarly, some constructed emotion researchers have noted that there may be emotion concepts shared across a broad range of human populations (Mesquita & Frijda, 1992). This consistency of concepts may result in some consistency in otherwise varying emotion instances.

Thus, even when agreeing about deviations from the absolute versions of their respective viewpoints (as described above), basic emotion and constructed emotion researchers differ on the underlying mechanics responsible for this deviation.

Do Emotions Generally Unfold Serially or in Parallel?

Consider how the various response types in an instance of "fear" may unfold. Some researchers have noted that such an instance typically arises with an initial appraisal of a situation (Ellsworth, 2013) and may involve a sequential unfolding of sub-component processes (e.g. Aue et al., 2007; Grandjean & Scherer, 2008). More generally, researchers with the appraisal-based perspective (Arnold, 1960; Lazarus, 1991; Smith & Ellsworth, 1985; Sander et al., 2005; Scherer, 1988) have emphasized that emotion instances often (but not necessarily always) feature an initial appraisal and other responses follow that initial appraisal.

Other researchers have noted that emotion instances may begin without an appraisal (Parkinson, 2007) and often involve interactive and reciprocal influences between response features. For example, researchers studying facial feedback (e.g., Coles et al., 2019) have suggested that appraisal-less facial expressions may initiate and profoundly affect emotional responding. Similarly, researchers have found that merely directing participants to assume facial expressions related to certain emotion categories (i.e. without appraisal) can result in autonomic responses that typically occur in emotion instances of that category (Ekman et al., 1983).

Towards an Integrative Framework

If one accepts that emotions are either consistent or variable, or that they are either appraisal-led and sequential or diversely led and interactive, one must reject a large corpus of empirical data that is not consistent with the 'correct' perspective. Another possibility, however, and the one we develop here, is that emotions are consistent in some circumstances and variable in others. Similarly, they unfold serially from appraisals in some circumstances, and not in others. On this view, the key question is not whether a particular perspective is correct or incorrect, but rather when (i.e. in what circumstances) a particular perspective is applicable and when it is not. Developing such a possibility requires an integrative framework that can identify the particular circumstances in which an emotion instance has some properties but not others.

To create such a framework, we draw upon the Parallel Distributed Processing (PDP) tradition (Rumelhart & McClelland, 1986), which features local computation, heterarchical processing, and the emergence of behavior. In particular, we rely upon the Interactive Activation and Competition (IAC) architecture (McClelland, 1981; McClelland & Rumelhart, 1981) that embodies a specific set of assumptions within the broader PDP tradition.

We build our argument over two sections. In Section 1, we describe the key principles of the IAC framework for Emotion (IAC-E) and describe the occurrence of an emotion instance in the IAC-E. In Section 2, we explore how emotion instances represented in the IAC may be consistent in some circumstances and variable in others and how they may be appraisal-initiated and sequential in some circumstances and input-agnostic and interactive/reciprocal in others. We use IAC-E based simulations to provide algorithmic explanations for several empirical phenomena related to emotion. We conclude by inviting consideration of the IAC-E as integrative and generative framework for emotion research.

The Interactive Activation and Competition Framework for Emotion (IAC-E)

The PDP tradition (Rumelhart et al., 1986), sometimes referred to as connectionism, seeks to explain mental processes using artificial neural networks. The IAC framework (McClelland & Rumelhart, 1981; McClelland, 1981) is a particular type of artificial neural network. In addition to its initial application to context effects in perception, the IAC framework has been used to examine perception (McClelland et al., 2014), emergent category formation and category-based inference (McClelland, 1981), social cognition (Freeman & Ambady, 2011), memory (Kumaran & McClelland, 2012), social behavior (Ehret et al., 2015; Read & Miller, 2002), legal judgments (Simon et al., 2015), emotional consciousness (Thagard, 2008), probabilistic inference (Glöckner et al., 2010), and value-based decision making (Suri et al., 2019). Here, we consider whether the IAC framework can be applied to emotion. We refer to our framework as the IAC framework for Emotion (IAC-E).

Key Principles of the IAC-E

Like other artificial neural networks, the IAC-E framework assumes that all processing occurs within neuron-like elements called units. These units influence each other via weighted connections. All connections in the IAC-E are bidirectional, and may be positive or negative.

All knowledge is resident within these weighted connections. Learning in the network occurs either by creating new connections between units or by updating existing connection weights.

The units (depicted by small circles in Figure 1) are organized into pools. There are two types of pools: input pools and hidden pools. Input pools can receive input from outside the network – however their final activation reflects both the input provided and the activation of other units that they are connected with. The activation of hidden units is completely determined by the activation of other units that they are connected with (they cannot receive external input). Positive weights tend to increase activation in the receiving unit, whereas negative tend to decrease activation in the receiving unit. Positive weights connect units belonging to different pools (i.e. input and hidden). Negative weights may exist within units of the same pool.

For illustrative purposes, we shall assume six input pools (denoted as parallelograms in Figure 1), however the number of pools is a free parameter in the framework (i.e. one may have any specified number of pools greater than one). The units of the input pool represent the features associated with an instance of an emotion. In Figure 1, for illustrative purposes, we use pools representing features related to cognition, motivation, behavior, neural activation in particular circuits, somatic physiology, and experience. We do not

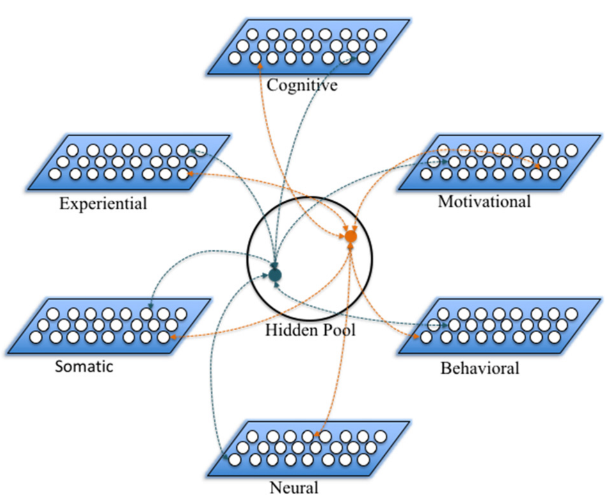

Figure 1. A representation of the interactive activation and competition framework for emotion (IAC-E). The parallelograms are input pools that represent feature sets of an emotion instance. Six pools are shown for illustrative purposes, but the framework is agnostic to which particular feature pools are used. The units are represented by white circles within each input pool and are organized by dimensions (2 dimensions – rows and columns – are shown, but any number of dimensions are possible). The circle in the middle is the hidden pool that contains two (conjunction) hidden units. Connections between hidden units and feature units may be innate or learned.

claim that this list – adapted from prior reviews on the nature of emotion (Moors, 2017) – is exhaustive, or that its components are mutually exclusive. It is also evident that the features we have listed are not all at a similar level of complexity or granularity. Different theorists may therefore prefer different lists of features, or may regard one or more of the input pools we include to be better represented as hidden pools rather than input pools (e.g., emotion experience). Our intent is not to suggest a definitive list of feature-sets. We merely seek to offer it as one example of what a list of emotion features might look like. The IAC framework for emotion is agnostic to the particular list of features that are included in the network.

Within each feature pool, the units are organized along one or more dimensions. For example, the ‘somatic’ pool might be composed of a dimension representing blood pressure and heart rate. Figure 1 features units organized along two dimensions (depicted via 3 rows and 8 columns), but in principle, a pool could have any number of dimensions. Once again, the IAC framework for emotion is agnostic to the particular dimensions that are included in each pool.

We next consider four core principles of the IAC-E. Similar principles are common to most IAC networks, but our description and examples emphasize their usage in contexts involving emotion.

The Connectivity Principle. Connections between units may be excitatory or inhibitory. Excitatory connections result in the amplification of activation in the involved units, whereas inhibitory connections lessen the activation of the involved units. In the IAC-E, all connections are bi-directional and equal. Bi-directional networks are guaranteed to converge (Perfetti, 1993) such that for any input, the network always attains a stable state. Connections between units in different pools are always excitatory and connections with units in the same pool are always inhibitory. These properties are summarized in the Connectivity Principle.

> Principle 1 – Connectivity: In the IAC-E connections between units in different pools are always excitatory, and connection with pools are always inhibitory. All connections are bidirectional and equal.

Inhibitory connections are a core element of mammalian brains (e.g. Kisvárday et al., 1997) and they are important enablers of competition between units in the IAC-E. Competition (when present) allows the network to represent and act upon the feature with the highest activation rather than responding to an array of potentially contradictory feature units. For example, assume that some elements in an environment cause approach-related motivation units to activate while others cause avoidance-related motivation units to activate. Further, assume that activation related to approach-related motivation units is higher than activation related to avoidance-related motivation units. If there were no inhibitory connections between the approach units and the avoid units, then the network’s response would be, at best, weakly oriented towards an overall approach response. If, however, the approach units and avoid units have inhibitory connections, then the slight lead of the approach units will suppress the less activated avoid units. In this case, due to the "rich get richer" phenomenon in IAC networks (McLeod et al., 2006), units corresponding to the approach motivation units (and not the avoid units) would be activated and the network would be robustly oriented towards an approach response.

The Input Principle. The network cannot receive direct input into the conjunction units in the hidden pool (hidden in the sense that this pool cannot receive external input). Only the feature units in the input pools can receive external input.

> Principle 2 – Input: External input into the network can only be provided into any feature units in any input pool (i.e. conjunction units in the hidden pool cannot directly receive external input). Conjunction units, once activated, can reciprocally influence any feature unit.

The input principle ensures input into emotion networks could potentially arise from any feature unit in any feature pool. Importantly, feature units in the input pools are not merely activated by external input. Rather, they could receive activation from conjunction units via reciprocal connections. Activation in feature units rises and falls during the course an emotion, and collectively these activation patterns represent an emotion instance.

The Interactivity Principle. The occurrence of simultaneous changes in different features is a central aspect of an instance of emotion. For example, an instance of fear may involve simultaneous changes in units representing peripheral physiology (in the somatic pool) and units representing avoidance tendencies (in the motivational pool), and other pools as well. It is unlikely such simultaneity is exclusively attributable to different features independently responding to an emotion-eliciting stimulus at the same time. This would require each feature to be exclusively responsive to emotion-eliciting stimuli and exclusively unresponsive to co-occurring features. It is more plausible that changes in some features – represented by activations of corresponding units in the IAC network – cause changes in other feature units via the influence of other features. In the IAC-E (like all neural networks), we capture such influences via a transference of activation. One possible way to ensure such transference to postulate direct connections between some feature units. Another option – and one we develop in the present work – is to postulate that feature units are connected to each other via units in the hidden pool.

> Principle 3 – Interactivity: Feature units in input pools are connected to each other via conjunction units in the hidden pool.

The interactivity principle entails that a feature unit that is activated by an emotion-eliciting stimulus would in turn activate the hidden units that it is connected with, and these activated hidden units would then activate other feature units that are connected with these hidden units. Thus, hidden units ensure interactivity between feature units in the IAC-E and enable network flexibility and context sensitivity. If feature units were directly connected with each other, the activation of one feature unit would reliably cause activation in the feature units it is connected with which would decrease the context sensitivity of the network.

The Learning Principle. Connections between feature units and hidden units may be innate or learned. Innate connections are assumed to be present from birth. Learned connections between feature units are acquired via experiences in which an emotion-eliciting stimulus causes simultaneous activations in different sets of feature units which are then connected to each other via sets of hidden units. Hidden units enabling such learned conjunctions between feature units become functional with experience and are not present from birth. Some connections may start out as being innate and may be transformed with experience.

> Principle 4 – Learning: Connection weights between conjunction units and feature units may be innate (present from birth) or learned. Learned connection weights are formed with experience when two or more feature units co-occur due to statistical regularities in the environment.

Importantly, some (innate or learned) connections may be numerous and (indirectly) connect similar sets of feature units. Collectively such populations will exert a strong influence on network dynamics. Other conjunction units that are not participating in large populations connecting similar sets of feature units will exert less influence on network dynamics.

The instance-based approach that we have taken here is an algorithmic claim about the functional behavior of our brains, and not a claim that the brain literally represents each individual emotion experience separately (in different units). Ideas about how distributed representations, in which a small set of units, each participating in the representation of many individual instances, could capture the properties of instance-based models were first introduced by Marr (1982) and have subsequently been explored by many authors as alternatives to the idea that we actually allocate individual neurons or distinct non-overlapping sets of neurons to individual items (overview provided in Plaut & McClelland, 2010).

Notably, some of the key principles of the IAC-E were anticipated by Gordon Bower and his colleagues (Bower, 1981, 1992; Bower & Forgas, 2000) in their Human Associative Memory (HAM) model. According to the HAM model, emotions could be represented as units within a network, and could form associations with coincident events (which were also represented by units). The spread of activation in the HAM model began with external input and was thought to be akin to the flow of water or electricity. The HAM model was able to explain various effects including mood-state-dependent memories, emotional/social judgements, and some types of categorizations. However, the links in HAM network models were treated in an ad-hoc manner and could have a range of labels (such as “has a”, “is not a”, or “name” etc.). Such unexplained flexibility proved difficult to justify, and was finally resolved via PDP models that featured interactive and hierarchical processing of the type included in the IAC-E.

An Instance of Emotion in the IAC-E

In the IAC-E, an emotion instance begins when one or more emotion-related feature units receive input from outside the network. As an example, consider the case in which input begins in the somatic pool (e.g. due to an infant experiencing discomfort). This feature unit may be connected to a behavior unit (e.g. crying) via a (conjunction) hidden unit. Input into the somatic unit produces activation in the unit in the hidden pool. Rising activation in the hidden pool unit then activates the feature unit in the behavioral pool. Activations from the behavioral unit (and the somatic unit) then reinforce activation in the hidden unit (and are reinforced by it). Over time, the input into the somatic feature unit may decrease and/or the activation of the units may lessen via an in-built decay function. This rise and fall of activation in the network – enabled by an external input into one or more emotion feature units, its interactive spread to other emotion feature units, and its decay back to baseline – constitutes an instance of emotion in the IAC-E.

Other instances of emotion may be far more complex. Consider the case of an adult walking into a job interview. Several emotion-related feature units may be relevant and they may have many associative links with each other via conjunctive hidden units.

The hidden units are illustratively shown as two clusters because all units within each cluster have prior connections to the same input units in the cognitive, somatic, and behavioral pools (these pools were chosen as illustrative pools; any other combination of pools could have been specified). For example, the light-colored cluster units may be connected with units representing pleasant anticipation at the prospect of getting the job. The cluster of darker units – fewer in number – might be connected with feature units representing anxiety related to being interviewed. The light and dark colored clusters are connected to different somatic and behavioral units.

It is instructive to analyze the dynamics of this more complex situation. Let’s say an event in the world occurs (e.g. the prospect of a job interview) and is represented by input in unit C18 (first row, 8th column with respect to the bottom-right corner). Input in this unit activates both collections of units in the hidden pool. These collections compete with each other (since they have inhibitory weights not shown in Figure 2). The overall activation of the darker collection is reduced in comparison to units in the lighter collection – because the latter has more units. Units in the lighter collection in turn activate units S33 and B11 in the somatic and behavioral pools respectively. Units in the darker collection in turn activate S21 and B31 – however these latter units are activated less strongly than activations in S33 and B11 since a greater number of hidden units were voting for them (i.e. sending activation to them). The units in each input pool (e.g. S33 and S21) are competitive with each other (i.e. they have inhibitory weights). As the interactivity builds, it displays a compromise level of activation (e.g. between S33 and S21 but weighted towards S33 since it received a greater activation flow). Finally, activation levels reduce (due to reduced levels of input and activation decay) and the emotion instance is completed.

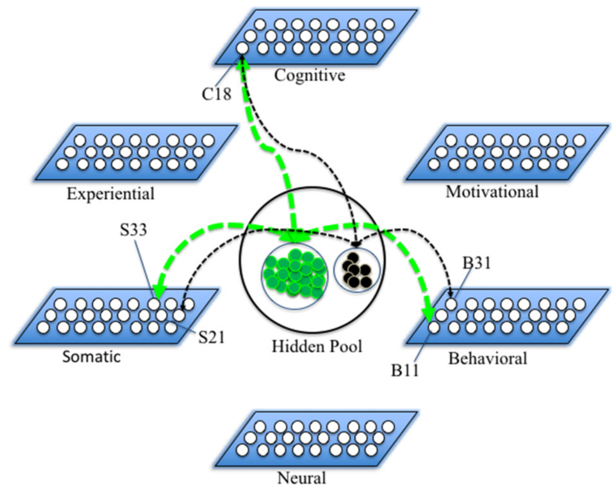

Figure 2. Illustrative network dynamics of an instance of an emotion in the IAC-E. This emotion instance begins in the cognitive pool with input into the unit labeled C18 (an emotion instance could begin with input into any feature unit in any pool). This input activates two populations of units in the hidden pool, which in turn activate competing units in the behavioral and somatic pool. See text for details.

The IAC-E Applies at Different Levels of Granularity

Thus far, we have worked with individual feature units that are connected to each other via individual conjunction units. All positive connection weights equal some small constant value (say $+\epsilon$), and all negative connection weights equal some small negative constant value (say $-\epsilon$). However, the principles of the IAC-E described above are equally applicable to collections of feature units. In this case, multiple feature units would be combined and represented as a single summary unit. The connection weight between two summary units is simply the sum of the individual weights of the units contained in the summary units. For example, if each of two summary units contain 5 individuals each, each of which are connected with a corresponding individual unit in the other summary unit, then the weight between the two summary units would be 5ϵ. In practice (for the simulations described in the next section), we shall cap the maximum positive weight between units at $+1$ and the maximum negative weights at -1. This is a common assumption in most IAC frameworks.

The granularity of the collection chosen for a particular summary unit depends on the goals of the modeler performing the simulation. For example, it is, in principle, possible to specify two feature units, each representing an electrodermal response of 2.241 and 2.242 Micro Siemens respectively. If an experimenter seeks to model such small differences, then her IAC-E network would feature separate units representing these quantities. However, if she is interested in more macro-properties (e.g. electrodermal responses of less than 2 vs. responses greater than 5) and not in small differences in emotion features, then she may collect many granular units into fewer summary units. This approach is in line with exemplar models used in many domains of psychology (e.g. Ashby & Rosedahl, 2017; Hintzman & Ludlam, 1980; Medin & Schaffer, 1978).

An individual feature unit represents a particular response feature of an instance of an emotion. A granular feature represents an approximate summary of a property relevant to an emotion instance. Similarly, an individual conjunction unit represents the summary of a single instance of an emotion. A granular conjunction unit represents an approximate summary of past experiences with a collection of instances of an emotion.

Types of Connections in the IAC-E

Thus far we have described how connections between units in the IAC-E may be innate or they may be learned. As discussed above, we have allowed for the possibility that each type of connection may have varying strengths (between -1 and $+1$, by convention).

We assume that it is possible to specify a set of emotion instances such that each one of them feature a high percentage of strong and innate connection weights. We also assume that it is possible to specify a set of emotion instances such that each one of them feature a high percentage of learned connection weights of varying strengths. This assumption does not specify a particular threshold for what constitutes a 'high percentage' because doing so requires observers to agree on a canonical list of emotion features.

In the absence of such an agreed-to canonical list, the spirit of our assumption is that both innate and strong connections as well as learned connections of varying strength are ubiquitous in what people generally refer to as instances of emotion.

The IAC-E is Falsifiable

As we have discussed above, The IAC-E is capacious enough to support emotion instances of different types without sacrificing its internal consistency. However, this does not mean that the IAC-E supports *any* pattern of emotion instances and cannot be disproved. Rather a number of empirical observations could falsify the IAC-E's core assumptions. For example, if most populations of neurons that are active in representing particular emotion features are directly connected to each other (i.e. not via conjunction neurons), then the interactivity assumption would not hold. In such a situation, context effects in emotion (which have often been empirically observed) would have to have a different neural mechanism other than the interactivity-based mechanisms offered by the IAC-E. Similarly, consistency of responses in the IAC-E is based on feature units that are strongly connected with other feature units. We have proposed that such connections must either be innate or developed via universal statistical regularities. Once again, this is a testable (and falsifiable) claim.

The potential generativity of the IAC-E stems from its implications. We turn to these next.

Implications of the IAC-E

We propose that the IAC-E has two important implications, each of which we discuss in the following two sections. These implications are related to (1) The IAC-E being an integrative framework for various instances of emotion; (2) The IAC-E enabling the quantitative simulation of several well-known empirical phenomena related to emotion.

The IAC-E as an Integrative Framework

We propose that under certain conditions the IAC-E can represent instances of emotions comprised of consistent responses to similar inputs and in other situations the IAC-E can represent instances of emotion comprised of variable responses that are shaped by experience and time. Additionally, under certain conditions, the IAC-E represents emotion instances that are not exclusively initiated by any single input pool (input agnosticism) and feature interactive and reciprocal effects. However, under certain (different) conditions, the IAC-E represents emotion instances that are appraisal-led and may proceed sequentially.

Representing Consistent as well as Variable Emotion Instances. The IAC-E enables consistency of emotion response patterns via at least two mechanisms. In both mechanisms, consistency of response patterns occurs due to the strong influence of a large number of conjunction units exerting their collective influence.

The first mechanism enabling consistency is the presence of a large population of hidden units that innately connect selected feature input units to each other. In such a situation if a feature pool unit (say unit A) receives external input, it activates the population of hidden units it is innately connected with; these units in turn activate the other feature units they are innately connected with (say units B and C). If there are enough such similarly-connected conjunction units they will collectively overcome the influence of (presumably fewer) context-related feature units. In such cases, activation in unit A will consistently produce activation in units B and C.

The second mechanism enabling consistency relies upon statistical regularities in the environment to produce a population of well-connected learned hidden units (postulated by the learning principle). Here the hidden units were not innate, but developed over time in consistently co-occurring populations of input units. Such units will produce similar consistency as innate units.

Related to variability, connecting feature units to each other via conjunctive hidden units (postulated by interactivity principle) affords the opportunity to explain the effects of experience and context of emotion-related responses over time. For example, consider the case of a person who has experienced many instances of emotions in which a higher than normal level of somatic arousal is associated with withdraw motivation (corresponding, for example, to fear) and only a few instances of emotions in which a higher than normal level of somatic arousal is associated with approach motivation (corresponding, for example, to excitement). If such a person experiences a high heart rate (an activation of units in the somatic input pool), she is likely to exhibit increased activation in withdraw motivation units. In the IAC framework for emotion, this occurs because the 'increased heart rate' input unit activates many prior hidden units – each of which send activation to the 'withdraw' motivation unit. Contrastingly, fewer hidden units send activation to the 'approach' motivation unit. Eventually the activation in the 'withdraw' feature unit will inhibit the activation in the 'approach' feature unit (postulated by the competitive feature unit principle).

However, this association between 'increased heart rate' and 'withdraw' motivation units can be modified by experience: imagine that this person next encounters several other instances of emotion that are associated with a high heart rate, but are also associated with an approach motivation (e.g. because she frequently rides on an enjoyable Ferris wheel). These emotion instances connect via new conjunction units in the hidden pool (due to learning). Now the 'high heart rate' feature units would activate conjunction units that send activation to 'approach' (as well as earlier conjunctions sending activation to 'withdraw'). This would effectively weaken the association between experiencing a high heart rate and wanting to withdraw. With enough new hidden unit conjunctions, a high heart rate may even become associated with approach motivation. Hidden units thus play the role of casting top down votes that determine the effective associative strength between different emotion features.

Variability may also arise due to contextual variables because associations involving hidden units are also subject to selective activation based on combinations of input cues (Medin & Schaffer, 1978; McClelland, 1981). Increasing specificity of input can alter the pattern of activated conjunctions in the hidden pool. Continuing with our example above, providing input into the 'high heart rate' and 'withdraw' feature units will activate hidden units corresponding to instances of fear and disgust; this activation may in turn activate the feature units associated with the subjective experience of fear or disgust. On the other hand, input into the 'high heart rate' and 'approach' feature units will activate the hidden unit corresponding to instances of excitement; this activation may in turn, activate the feature units associated with the subjective experience of excitement.

Representing Input-Agnostic as well as Appraisal-led Emotion Instances. The input principle postulates that any feature unit in any input pool may receive external input. Thus, in principle, emotion instances need not be initiated by activation in a single pool (e.g. the appraisal pool). Once a set of feature units activate a population of conjunction units, those units may activate a different set of feature units, which may dynamically influence the original set of feature units. Thus, the IAC-E supports interactive and reciprocal influences in the unfolding of an emotion instance.

While the IAC-E framework has no requirements that emotions begin with external input in the appraisal pool, it supports the possibility that many emotion instances begin with an appraisal. Further while interactivity and reciprocity are modal in the IAC-E, it supports the possibility of observing emotion instances that are appraisal-led and appear to unfold sequentially. For example, consider the case in which one feature unit, say F1, receives external activation and activates a set of conjunction units, say C1, that are connected with a feature unit say F2. Once F2 is activated, it may activate a different set of conjunction units say C2, which may then activate a third set of feature units F3. Such an activation pattern will cause a sequential spread from F1 to F2 to F3.

More generally, appraisal led emotion instances require a large number of strongly connected units in the appraisal feature pool. Such units may arise with repeated input into one or more appraisal units – which may, for example, be the case for human adults. Such repeated external input will cause strengthened connections between the appraisal feature pool and the conjunction pool (due to the learning principle).

Simulations of Empirical Findings Using the IAC Framework

In this section, we seek to test whether the IAC-E can simulate well known empirical findings from the emotion literature. We reasoned that such simulations, if successful, could strengthen the claim for the IAC-E to be a general framework for emotion.

Our goal in these simulations is to test the consequences of the principles of the IAC-E. If diverse empirical phenomena in the emotion literature can be explained as logical consequences of these principled (and only these principles), then one must increase one's confidence that the IAC-E offers a useful algorithmic account for emotion phenomena. We shall detail two simulations below to illustrate the IAC-E model. Further simulations are described in the Supplementary Materials.

The first simulation is concerned with the influence of context on emotion perception. The second simulation is concerned with consistency of emotion responses across people. Other simulations, using similar techniques as those showcased below, related to the facial-feedback effect (Noah et al., 2018) and how emotion words shape emotion percepts (Gendron et al., 2012), are described in the Supplementary Materials. The particular feature units chosen for the simulation, while always obeying the IAC-E assumptions, depend on the particular empirical phenomena being modeled. All specific algorithmic details of the IAC-E network are identical to other interactive activation networks and have been described elsewhere (Suri et al., 2019).

Simulation #1: Malleability of Emotion Perception. A large number of empirical studies (e.g. Hess et al., 2016; Matsumoto & Hwang, 2013) have examined the effects of context (e.g. making a fist) on facial emotion recognition (e.g. recognizing that a face depicts the emotion of anger). This simulation proposes an IAC-based mechanism for such effects. Here we take the perspective of an emotion observer estimating emotion dynamic using her own IAC-E.

In a well-known study, Aviezer et al. (2008) presented participants with images of an individual with various facial expression/context combinations and asked to choose the emotion that "best describes the facial expression" from a list of six basic-emotion labels (disgust, anger, sadness, fear, surprise, and happiness). The stimulus set consisted of posed disgust faces superimposed onto models in four emotional contexts: disgust, anger, fear, and sadness. Context, here, refers to all information external to the face, including body posture and the surrounding scene. For example, the disgust context depicted the person holding a soiled diaper, and the anger context depicted the person making a fist. Importantly, while the context variables varied across images, the face did not (i.e. it was the same face in all images – always showing a disgust face).

Categorization of facial expressions matched the context emotion most often when the emotional context matched the facial expression typically associated with such context (i.e., a disgusted expression in a disgust-suggestive context; see empirical results in Figure 3a and 3b for details). For example, in a disgust context, respondents accurately classified the face as representing disgust in 91% of trails. However, in an anger context, respondents only classified the face representing disgust in 11% of trials (as depicted in Figure 3a); in 87% of trials, respondents were influenced by context and classified the face as representing anger (as depicted in Figure 3b). The effects of context were largest for anger, second for sadness, and weakest for fear.

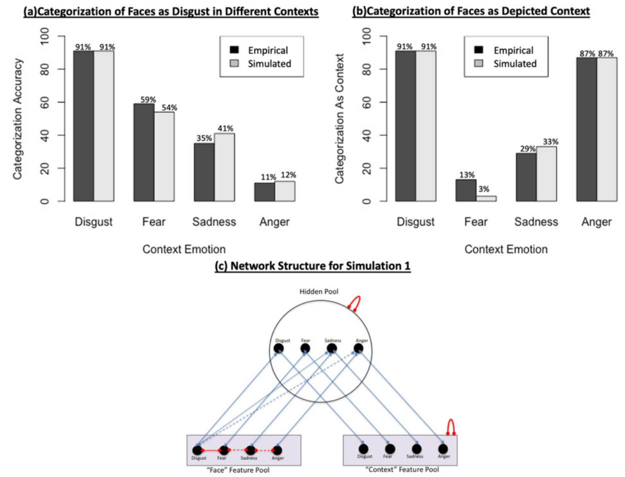

Figure 3. Empirical and simulation results for simulation 1. Panel (a) shows empirical data and simulation results for categorization accuracy in different contexts. Here a disgust face in a disgust context is accurately classified in most (91%) of trials, but a disgust face in an anger context is infrequently classified as disgust. Panel (b) shows the empirical data and simulation results of the frequency with the disgust face is labeled with the context emotion. For example, in the anger context, the disgust face is labeled as anger in 87% of trials. Panel (c) shows the network structure used in the simulation.

Figure 3c represents the network structure used in the experiment. There are two feature pools – one for the face and one for the context. Both feature pools contain units for disgust, fear, sadness, and anger – the face feature pool receives input from face-related features, and the context feature pool receives inputs from context-related features. The hidden pool (depicted as a circle in Figure 3c) contains an integrative representation of disgust, fear, sadness, and anger. The units of the hidden unit receive activation from – and provide activation to – units in the face feature pool and context feature pools.

Since integrative representations of the emotion categories in this empirical context are well differentiated from each other, connection weights between units in the hidden pool are fully inhibitory (i.e. the connection strength between each hidden unit is −1). This is diagramed using the loop on the top right of the hidden pool in Figure 3c. Similarly, since contexts depicted in the experiment (e.g. a diaper for disgust and a fist for anger) are well-differentiated connection weights between units in context feature pool are also fully inhibitory (i.e. the connection strength between each unit in the context feature pool is −1). Units in the face feature pool are unevenly differentiated from each other – some units share common features whereas others are highly differentiated. Following a computational model developed for assessing similarities between facial expressions (Susskind et al., 2008) facial expressions of disgust are highly differentiated from facial expressions of fear, somewhat distinct from facial expressions of sadness, and relatively similar to facial expressions of anger. Constrained by this model, we assigned an inhibitory weight of −1 between "face" feature units of disgust and fear, a weight of 0.8 between disgust and sadness, and a weight of −0.5 between disgust and anger.

There were excitatory weights (+1) between units in the context feature pool and corresponding units in the hidden pool. For example, the 'sadness' context unit connected with the 'sadness' hidden unit, but not to any other hidden units. Facial feature units were also connected with their corresponding feature units (+1 connection weight). However, they could also be connected with other hidden units that they were somewhat consistent with. For example, the disgust facial unit was connected to the disgust hidden unit with +1 weight, with the fear unit with 0 weight (i.e. unconnected) with the sadness hidden unit with +0.2 weight, and with the anger hidden unit with +0.5 weight. These weights reflected the similarity between a facial feature and an emotion (Susskind et al., 2008).

In all conditions, the disgust facial feature unit was provided with an input of 1. The context units were provided activation depending on the context being modeled. In the disgust context, the disgust context unit was given input equal to 1, in the anger context, the anger context unit was given input equal to 1 (and similarly for the sadness and fear contexts). The input activation from the feature units flowed to the hidden units, and then back to the feature units (since all weights in the network are bi-directional) until the network converged. A Softmax function ($\lambda = 7$) was used to calculate the probability that a face was classified as disgust in each condition (Figure 3a) and the probability that the face was classified in line with the context emotion (Figure 3b).

This simulation showcases how context effects might change the perception of emotion faces. Activation was the only currency in the network – it did not rely on affect programs to instantiate emotions, nor was it solely dependent on conceptual knowledge related to emotion contexts. It showcased a mechanism which allowed context to affect faces and also made the prediction that facial expressions would impact the perception of ambiguous contexts.

Simulation #2: Consistent Autonomic Activity in Response to Consistent Input. In this simulation, we propose a mechanistic explanation of how autonomic activity produced via facial prototypes or via reliving past experiences displays consistent footprints in response to consistent network input the categories of anger, fear, and disgust.

Ekman et al. (1983) measured autonomic activity in two tasks. In the directed facial action task, in a series of trials, participants were asked to construct facial prototypes of emotion by arranging facial muscles into an emotion-prototypic expression. They were not required to produce an emotional expression but instead were told which facial muscles they should contract. Video-records were used to only include those trials in which participants successfully followed instructions. In the relived emotion task, in a series of trials, participants were asked to experience emotions by reliving a past emotional episode. Participants rated the intensity of felt emotion after every trial, and only those trials in which their reported intensity was greater than the mid-point, were used for analysis.

The experimenters reported findings in two autonomic categories: changes in heart rate, and changes in (right) finger temperature. They found consistent differences in the emotion categories of fear, anger, disgust, surprise, happiness, and sadness. For simplicity, in this simulation, we illustrated the first three emotion categories only (an identical mechanism can be posited for the other three), and we therefore report on the results for those categories only. Across both tasks, anger and fear were associated with higher heart rates, and disgust was associated with lower heart rates. Additionally, across both tasks, anger was associated with a higher temperature, whereas fear and disgust were associated with lower temperatures.

The network for this simulation (Figure 4) consists of a Face feature pool and a Relived emotion feature pool – each of which have units corresponding to the emotion categories of disgust, fear, and anger. There is a Hidden pool representing summary units for the three categories. Finally, there were feature pools containing units for higher or lower heart rates and higher or lower temperatures.

Figure 4 contains the connection strengths for Simulation 2. The connections involving the Face feature pool and the

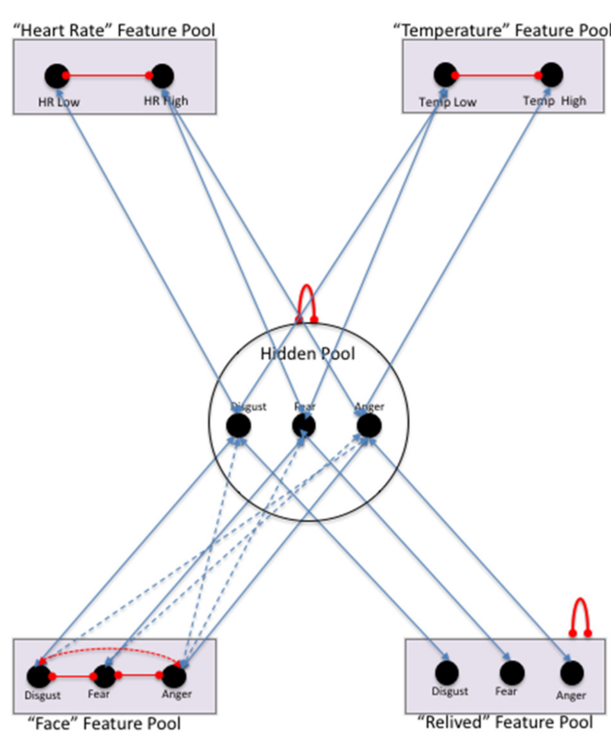

Figure 4. Network structure and activation dynamics for simulation 2.

Table 1. Activation values for input into face feature units in simulation 2.

	Low HR	*High HR*	*Low Temperature*	*High Temperature*
Disgust	0.31	−0.12	0.31	−0.12
Fear	−0.13	0.41	0.30	−0.10
Anger	−0.13	0.35	−0.13	0.35

Hidden pool are constrained with those described in Simulation 1. All other connections are fully excitatory (+1, between pools) or fully inhibitory (−1, within pools).

We simulated 6 conditions: in the first three conditions we provided +1 input, respectively, to the disgust, fear, and anger face feature units. In the next three conditions, we provided +1 input, respectively, to the disgust, fear, and anger relived emotion feature units. In each condition, the feature unit activated the hidden unit it was connected, which in turn activated the appropriate feature heart rate and temperature feature units. As expected, activation values upon convergence were very similar in Conditions 1 and 4, 2 and 5, and 3 and 6. The values for Condition 1 (input into Disgust Face unit), Condition 2 (input into Fear Face unit), and Condition 3 (input into Anger Face unit) are shown in Table 1.

The IAC-E can display consistent response patterns in response to consistent network inputs. Further, this simulation is an instance of an emotion that does not begin with an appraisal (rather, some conditions are initiated via posed facial expression causing activation in the face feature pool).

Concluding Comment

The philosopher of science, Thomas Kuhn, is said to have noted that "the answers you get depend upon the questions you ask" (as cited in Shea, 2004). A parallel point can be made related to scientific research in general, and emotion research in particular: the emotion perspectives we develop depend upon the empirical phenomena we use to inform them.

For example, some researchers have attended to and been informed by empirical phenomena that showcase the consistency of emotion responses, and have concluded that emotions are generally consistent responses to recurring situations. Other researchers have attended to and been informed by empirical phenomena that have showcase the malleability of emotions, and have concluded that emotions are profoundly shaped by context and learning. Similarly, some researchers have attended to and been informed by empirical phenomena that often seem to be initiated by appraisal and have concluded that emotions are generally initiated by appraisals, whereas other researchers have emphasized that emotions may be initiated by a variety of inputs that reciprocally interact with each other.

At first, it appears that these perspectives of emotion are contradictory with each other. That is, it seems that emotions are either generally consistent or generally variable. Similarly, it seems that emotions either generally begin with appraisal or they don't. Debates about such issues currently dominate emotion research and are sometimes said to have led to "a hundred years emotion war" (Lench et al., 2011).

The core proposition offered by the IAC-E is that these perspectives need not be contradictory with each other. The IAC-E offers an integrated framework in which emotions are – under some model parameter values – often consistent, and under other model parameter values, they are often variable. More specifically, when the percentage of strong innate connections in the set of emotion instances being observed is high, emotions are likely to show consistency. Conversely, when the percentage of strong innate connections in the set of emotion instances being observed is low, emotions are likely to show variability across people and time.

Similarly, if a network has units in its appraisal pool that have received repeated prior external input from the environment, then their connection weights would be stronger, and they would be likely to represent appraisal-led emotions. An absence of such well-connected units will lead to input agnostic emotion instances with interactive influences.

Thus, the IAC-E can represent emotion instances of various types using a single set of principles. This suggests that it may be possible to model every empirical observation related to emotion using the IAC-E (we have provided four representative simulations, two in the main text and two in the supplemental material). Such an approach could allow experimenters to specify actual weights that connect features of different types. Over time, this could result in a catalog of specified feature weights. In the best case, these feature weights will be consistent across experiments and will find supporting neural evidence. Alternatively, such cataloging efforts may lead to the conclusion that one or more principles or assumptions of the IAC-E must be refined or replaced. Either eventuality points to a potentially generative role for the IAC-E.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Gaurav Suri https://orcid.org/0000-0002-0423-060X

Supplemental material

Supplemental material for this article is available online.

References

Arnold, M. B. (1960). *Emotion and personality*. Columbia University Press.

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124,* 472. https://doi.org/10.1037/rev0000064

Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, *74,* 347–357. *https://doi.org/10.1016/j.biopsycho.2006.09.001*

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38,* 1145–1160. *https://doi.org/10.1037/0003-066X.38.11.1145*

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., & Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19,* 724–732. *https://doi.org/10.1111/j.1467-9280.2008.02148.x*

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10,* 20–46. *https://doi.org/10.1207/s15327957pspr1001_2*

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36,* 129–148. *https://doi.org/10.1037/0003-066X.36.2.129*

Bower, G. H. (1992). How might emotions affect learning. In S. A. Christianson (Ed.), *Handbook of emotion and memory* (pp. 3–31). Erlbaum.

Bower, G. H., & Forgas, J. P. (2000). Affect, memory, and social cognition. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, & P. M. Niedenthal (Eds.), *Cognition and emotion* (pp. 87–168*). Oxford University Press.*

Coles, N., Larsen, J., & Lench, H. (2019). A meta-analysis of the facial feedback hypothesis literature: Effects of facial feedback on emotion experience are small and variable. *Psychological Bulletin*, *145,* 610–651. *https://doi.org/10.1037/bul0000194*

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114,* E7900–E7909. *https://doi.org/10.1073/pnas.1702247114*

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145,* 830–843. *https://doi.org/10.1037/xge0000172*

Ehret, P. J., Monroe, B. M., & Read, S. J. (2015). Modeling the dynamics of evaluation: A multilevel neural network implementation of the iterative reprocessing model. *Personality and Social Psychology Review*, *19,* 148–176. *https://doi.org/10.1177/1088868314544221*

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6,* 169–200. *https://doi.org/10.1080/02699939208411068*

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, *3,* 364–370. *https://doi.org/10.1177/1754073911410740*

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science (New York, N.Y.)*, *221,* 1208–1210. *https://doi.org/10.1126/science.6612338*

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, *5,* 125–131. *https://doi.org/10.1177/1754073912463617*

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118,* 247–279. *https://doi.org/10.1037/a0022327*

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21,* 500–508. *https://doi.org/10.1038/mp.2015.88*

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion (Washington, D.C.)*, *12,* 314–325. *https://doi.org/10.1037/a0026007*

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23,* 439–462. *https://doi.org/10.1002/bdm.668*

Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion (Washington, D.C.)*, *8,* 341–351. *https://doi.org/10.1037/1528-3542.8.3.341*

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science*, *71,* 901–911. *https://doi.org/10.1086/425944*

Hess, U., Blaison, C., & Kafetsios, K. (2016). Judging facial emotion expressions in context: The influence of culture and self-construal orientation. *Journal of Nonverbal Behavior*, *40,* 55–64. *https://doi.org/10.1007/s10919-015-0223-7*

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition*, *8,* 378–382. *https://doi.org/10.3758/BF03198278*

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2,* 260–280. *https://doi.org/10.1111/j.1745-6916.2007.00044.x*

James, W. (1884). What is an emotion? *Mind; A Quarterly Review of Psychology and Philosophy*, *9,* 188–205. *https://doi.org/10.1093/mind/os-IX.34.188*

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion*, *13,* 467–480. *https://doi.org/10.1080/026999399379140*

Kisvárday, Z. F., Toth, E., Rausch, M., & Eysel, U. T. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cerebral Cortex*, *7,* 605–618. *https://doi.org/10.1093/cercor/7.7.605*

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119,* 573–616. *https://doi.org/10.1037/a0028681*

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46,* 352–367. *https://doi.org/10.1037/0003-066X.46.4.352*

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137,* 834–855. *https://doi.org/10.1037/a0024244*

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, *5,* 356–368. *https://doi.org/10.1177/1754073913489750*

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science*, *19,* 898–903. *https://doi.org/10.1111/j.1467-9280.2008.02174.x*

Marr, D. (1982). *Vision*. Freeman.

Matsumoto, D., & Hwang, H. C. (2013). Assessing cross-cultural competence: A review of available tests. *Journal of Cross-Cultural Psychology*, *44,* 849–873. *https://doi.org/10.1177/0022022113492891*

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion (Washington, D.C.)*, *5,* 175–190. *https://doi.org/10.1037/1528-3542.5.2.175*

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112,* 179–204. *https://doi.org/10.1037/0033-2909.112.2.179*

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. In *Proceedings of the third annual Conference of the cognitive science society, Berkeley, California* (pp. 170–172).

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38,* 1139–1189. *https://doi.org/10.1111/cogs.12146*

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88,* 375–407. *https://doi.org/10.1037/0033-295X.88.5.375*

McLeod, P., Plunkett, K., & Rolls, E. T. (2006). The attraction of parallel distributed processing for modelling cognition. In J. L. Bermudez (Ed.), *Philosophy of psychology* (pp. 182–202*). Routledge.*

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85,* 207–238. *https://doi.org/10.1037/0033-295X.85.3.207*

Moors, A. (2017). Integration of two skeptical emotion theories: Dimensional appraisal theory and Russell's psychological construction theory. *Psychological Inquiry*, *28,* 1–19. *https://doi.org/10.1080/1047840X.2017.1235900*

Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology*, *114,* 657–664. *https://doi.org/10.1037/pspa0000121*

Öhman, A., & Wiens, S. (2003). On the automaticity of autonomic responses in emotion: An evolutionary perspective. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science. Handbook of affective sciences* (pp. 256–275*). Oxford University Press.*

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science*, *2,* 281–296. *https://doi.org/10.1111/j.1745-6916.2007.00045.x*

Parkinson, B. (2007). Getting from situations to emotions: Appraisal and other routes. *Emotion (Washington, D.C.)*, *7,* 21–25. *https://doi.org/10.1037/1528-3542.7.1.21*

Perfetti, R. (1993). On the convergence of reciprocal discrete-time cellular neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *40,* 286–287. *https://doi.org/10.1109/81.224306*

Plaut, D. C., & McClelland, J. L. (2010). Locating object knowledge in the brain: Comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychological Review*, *117,* 284–288. *https://doi.org/10.1037/a0017101*

Read, S. J., & Miller, L. C. (2002). Virtual personalities: A neural network model of personality. *Personality and Social Psychology Review*, 6, 357–369.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, *1,* 26.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP Models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Experiments in the microstructure of cognition* (Vol. 1). MIT Press.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110,* 426–450. *https://doi.org/10.1037/0033-2909.110.3.426*

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110,* 145. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2005). Emotion in human consciousness is built on core affect. *Journal of Consciousness Studies*, *12,* 26–42.

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion*, *23,* 1259–1283. *https://doi.org/10.1080/02699930902809375*

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18,* 317–352. *https://doi.org/10.1016/j.neunet.2005.03.001*

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126*). Springer, Dordrecht.*

Shea, S. C. (2004). The delicate art of eliciting suicidal ideation. *Psychiatric Annals*, *34,* 385–400. *https://doi.org/10.3928/0048-5713-20040501-19*

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109,* 369–394. *https://doi.org/10.1037/pspa0000029*

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48,* 813–838. *https://doi.org/10.1037/0022-3514.48.4.813*

Suri, G., Gross, J. J., & McClelland, J. L. (2019). Value-based decision making: An interactive activation perspective. *Psychological Review*, *27,* 153–185. *https://doi.org/10.1037/rev0000164*

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11,* 843. https://doi.org/10.1038/nn.2138

Thagard, P. (2008). *Hot thought: Mechanisms and applications of emotional cognition*. Bradford Book, MIT Press.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3,* 397–405. *https://doi.org/10.1177/1754073911410747*

Tracy, J. L., & Robins, R. W. (2004). Show your pride: evidence for a discrete emotion expression. *Psychological Science*, *15,* 194–197. *https://doi.org/10.1111/j.0956-7976.2004.01503008.x*